

Implementasi *K-Means Clustering* Pada Pemilihan Santri dan Santriwati Berprestasi

DOI: <http://dx.doi.org/10.35889/jutisi.v14i2.3150>

Creative Commons License 4.0 (CC BY – NC) 

Maimanah Salsabila Nasution^{1*}, Raissa Amanda Putri²

Sistem Informasi, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

*e-mail *Corresponding Author*: maysalsaa1605@gmail.com

Abstract

Managing large amounts of data on male and female students poses a challenge for Pesantren Darul Arafah Raya in assessing the academic performance of its students. The K-Means algorithm is one of the most widely used methods in data mining for clustering large amounts of data. This study aims to cluster students into two categories: high-performing and low-performing, using the K-Means Clustering method. The variables analyzed include academic grades, attendance, attitude, involvement in organizations/extracurricular activities, and academic history. Data from 1,010 students were processed, and the algorithm stopped at the 35th iteration with significant centroid centers. The clustering results showed that 490 students were classified as high-achieving and 520 as low-achieving. Performance validation using a Silhouette Coefficient of 0.63 indicates that the clustering quality is in the good category, so the grouping results can be considered optimal. Integrating K-Means into a web-based information system facilitates strategic decision-making for the boarding school and supports the development of specialized training programs to cultivate students' potential in a more targeted manner.

Keywords: *Data Mining; K-Means Clustering; Outstanding Santri; Education; Website*

Abstrak

Pengelolaan data santri/wati dalam jumlah besar menjadi tantangan bagi Pesantren Darul Arafah Raya dalam menilai prestasi akademik santri. Algoritma *K-Means* merupakan salah satu metode dalam data mining yang banyak digunakan untuk mengelompokkan data dalam jumlah besar melalui pendekatan *clustering*. Penelitian ini bertujuan untuk mengelompokkan santri dan santriwati ke dalam dua kategori, yaitu berprestasi dan tidak berprestasi, dengan memanfaatkan metode *K-Means Clustering*. Variabel yang dianalisis meliputi nilai akademik, absensi, sikap, keterlibatan berorganisasi/ekstrakurikuler, dan riwayat prestasi. Data sebanyak 1.010 santri dan santriwati diproses, dan algoritma berhenti pada iterasi ke-35 dengan pusat *centroid* yang signifikan. Hasil klasterisasi menunjukkan 490 santri dan santriwati tergolong berprestasi dan 520 lainnya tidak berprestasi. Validasi performa menggunakan *Silhouette Coefficient* sebesar 0,63 menunjukkan kualitas klasterisasi berada pada kategori baik, sehingga hasil pengelompokan dapat dianggap optimal. Integrasi *K-Means* ke dalam sistem informasi berbasis web memudahkan pihak pesantren dalam pengambilan keputusan strategis, serta mendukung penyusunan program pembinaan khusus untuk mengembangkan potensi santri secara lebih terarah.

Kata kunci: *Data Mining; K-Means Clustering; Santri Berprestasi; Pendidikan; Website*

1. Pendahuluan

Kemajuan teknologi informasi memberikan kontribusi yang signifikan dalam bidang pendidikan, khususnya pada aspek pengelolaan data yang semakin kompleks seiring dengan meningkatnya jumlah peserta didik [1]. Pesantren sebagai lembaga pendidikan juga menghadapi tantangan dalam menilai prestasi santri secara objektif dan efisien ketika data yang dikelola berjumlah besar [2]. Oleh sebab itu, penggunaan metode data mining melalui algoritma *K-Means Clustering*, menjadi relevan untuk mengidentifikasi dan mengelompokkan santri berprestasi [3]. Penggunaan metode ini tidak hanya membantu mempercepat proses analisis, tetapi juga

mendukung pengambilan keputusan strategis bagi pihak pesantren dalam merancang program pembinaan yang lebih tepat sasaran serta meningkatkan mutu pendidikan secara keseluruhan.

Pesantren Darularafah Raya terletak di Kec. Kutalimbaru, Deli Serdang, yang memiliki jarak tempuh sejauh 25 km dari pusat kota Medan. Pesantren Darul Arafah Raya, yang berlokasi di Kutalimbaru, Deli Serdang, memiliki kapasitas kelas dan fasilitas yang memadai, termasuk laboratorium komputer dan ruang belajar yang lengkap, serta jumlah tenaga pengajar yang mencukupi. Namun, akumulasi data santri/wati yang terus meningkat setiap tahunnya membuat evaluasi prestasi akademik dan keterlibatan siswa menjadi kurang terstruktur. Saat ini, penentuan siswa berprestasi hanya dilakukan melalui evaluasi nilai akademik dan kehadiran, sehingga kurang efektif dalam mengelola data dalam jumlah besar [4].

Sebagai upaya mengatasi permasalahan, digunakan algoritma *K-Means Clustering*, salah satu metode data mining yang mampu mengelompokkan data berskala besar maupun data yang saling tumpang tindih [5]. Metode ini memungkinkan identifikasi pola prestasi siswa berdasarkan nilai akademik dan tingkat keaktifan, sehingga memberikan dasar yang lebih objektif bagi pengambilan keputusan [6]. Peningkatan dalam pemrosesan data menuntut penerapan strategi dan metode yang sesuai dalam melakukan proses pembuatan kebijakan, salah satu metode yang tepat ialah dengan menggunakan metode *K-Means Clustering* [7].

Penelitian ini bertujuan untuk menerapkan algoritma *K-Means Clustering* dalam pengelompokan santri/wati Pesantren Darul Arafah Raya ke dalam kategori berprestasi dan tidak berprestasi. Tujuan utama penelitian adalah menghasilkan model klasifikasi yang mampu menyajikan informasi terstruktur mengenai tingkat prestasi santri, sehingga dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan akademik. Penelitian ini diharapkan mendukung peningkatan mutu pendidikan melalui perumusan strategi pembinaan yang lebih terarah serta perencanaan program pengembangan potensi santri secara berkelanjutan.

2. Tinjauan Pustaka

Penelitian sebelumnya yang berjudul *Data Mining Seleksi Siswa Berprestasi Untuk Menentukan Kelas Unggulan Menggunakan Metode K-Means Clustering* (Studi Kasus: MTS Darul Fikri). Penelitian ini bertujuan untuk menjawab tantangan dalam menentukan pengelompokan kelas unggulan di MTs Darul Fikri Bringin, Ponorogo, yang mengalami peningkatan jumlah pendaftar setiap tahunnya. Untuk mengatasi hal tersebut, diterapkan metode data mining dengan pendekatan Algoritma *K-Means Clustering* guna membantu mengelompokkan siswa berdasarkan kemampuan mereka. Data yang dianalisis mencakup nilai rapor, tingkat keaktifan, dan kehadiran siswa kelas VII. Pengolahan data dilakukan menggunakan RStudio, dengan memanfaatkan antarmuka RShiny dan bahasa pemrograman R. Hasil penelitian menunjukkan bahwa dari 100 data siswa, kelas unggulan (klaster 1) terdiri dari 23 siswa yang ditempatkan di kelas A. Sementara itu, kelas B (klaster 2) yang merupakan kelas reguler memiliki 25 siswa, dan kelas C serta D (klaster 3) yang juga termasuk kelas reguler memiliki total 52 siswa. Dengan penerapan metode ini, MTs Darul Fikri Bringin dapat lebih efektif dalam menentukan dan menyusun kelas unggulan sesuai dengan kemampuan siswa [8].

Penelitian terdahulu yang berjudul *Penilaian Pengelompokan Data Prestasi Siswa Dengan Menggunakan Metode K-Means Untuk Mengenal Siswa Berprestasi*. Penelitian ini dilaksanakan di Mts Hidayatul Ulumiyah Ujung Kubu dengan fokus pada upaya sekolah dalam meningkatkan prestasi akademik siswa sesuai standar pendidikan nasional. Besarnya jumlah data siswa menyebabkan proses identifikasi siswa berprestasi menjadi kompleks, karena harus mempertimbangkan aspek penguasaan materi (teori maupun praktik), nilai sikap, dan kehadiran. Untuk mengatasi hal tersebut, digunakan metode *K-Means Clustering* dalam pengolahan data siswa. Hasil penelitian menghasilkan tiga klaster dengan nilai *silhouette score* sebesar 0,41, yaitu klaster 0 berisi 80 siswa dengan nilai terendah, klaster 1 berisi 49 siswa dengan nilai tertinggi, dan klaster 2 berisi 82 siswa dengan nilai terbaik. Temuan ini membuktikan bahwa penerapan algoritma K-Means dapat membantu sekolah dalam mengelompokkan siswa secara lebih objektif sehingga mendukung pengambilan keputusan terkait peningkatan prestasi akademik.

Penelitian terdahulu di SMK TI Panca Dharma Stungkit, Kecamatan Wampu, Kabupaten Langkat, menyoroti peran penting pendidikan dalam meningkatkan kualitas hidup serta menjadikan mutu pendidikan sebagai prioritas pembangunan nasional. Penilaian prestasi siswa dilakukan berdasarkan aspek teori, praktik, keterampilan, sikap, kehadiran, serta keterlibatan dalam kegiatan ekstrakurikuler. Pengolahan data dilakukan agar menghasilkan informasi yang dapat dijadikan acuan dalam pengambilan keputusan, penelitian ini menerapkan metode K-

Means Clustering dengan bantuan perangkat lunak Tanagra 1.4.50. Dari 24 sampel data siswa, metode ini berhasil membentuk dua cluster, yaitu 12 siswa dalam kelompok berprestasi dan 12 siswa dalam kelompok tidak berprestasi. Hasil penelitian ini menunjukkan bahwa penggunaan algoritma *K-Means* dengan *software* Tanagra mampu memberikan gambaran yang jelas mengenai pengelompokan prestasi siswa, sehingga dapat mendukung sekolah dalam menentukan kebijakan peningkatan mutu pendidikan [9].

Berdasarkan kajian terhadap penelitian-penelitian terdahulu, metode *K-Means Clustering* telah banyak digunakan untuk mengelompokkan siswa berprestasi dengan memanfaatkan berbagai variabel, seperti nilai rapor, sikap, kehadiran, dan keterlibatan ekstrakurikuler. Penelitian sebelumnya umumnya menggunakan perangkat lunak seperti RStudio dan Tanagra, namun masih memiliki keterbatasan berupa jumlah data yang relatif kecil serta belum terintegrasi dengan sistem berbasis web, sehingga hasil klasterisasi hanya bersifat analitis dan belum dapat secara langsung mendukung pengambilan keputusan. Kebaruan penelitian ini terletak pada implementasi *K-Means Clustering* pada skala data yang lebih besar (1.010 santri dan santriwati) dan integrasinya ke dalam aplikasi berbasis web, kemudian penilaian santri dan santriwati didasari oleh 5 kriteria, yang berbeda dari penelitian sebelumnya, sehingga hasil klasterisasi tidak hanya menyajikan pengelompokan santri berprestasi dan tidak berprestasi, tetapi juga berfungsi sebagai sistem pendukung keputusan yang operasional dan strategis untuk merancang program pembinaan yang lebih terarah di Pesantren Darul Arafah Raya.

3. Metodologi

Metode kualitatif digunakan pada penelitian ini, metode ini digunakan agar dapat memperoleh pemahaman mendalam tentang kebutuhan, persepsi, dan tantangan yang dihadapi dalam pengelolaan data [2]. Dengan menggunakan teknik seperti wawancara mendalam, observasi, dan studi pustaka, metode ini memberikan wawasan penting untuk pengembangan sistem informasi yang efektif dan efisien dalam pemilihan santri/wati berprestasi. Penelitian ini dilaksanakan pada bulan Oktober sampai awal bulan November yang dilakukan pada tahun 2024. Penelitian dilakukan di Pesantren DA Raya. Pesantren Darul Arafah Raya terletak di Kec. Kutalimbaru, Deli Serdang, yang menempuh jarak sejauh 25 km dari pusat kota Medan.

3.1 Algoritma *K-Means*

K-Means ialah teknik *Clustering* sederhana yang dikenal mampu bekerja dengan cepat dan efektif dalam mengelompokkan data. *K-Means clustering* merupakan teknik pengelompokan data non-hierarki yang digunakan untuk memisahkan sekelompok data menjadi beberapa klaster berdasarkan kemiripan karakteristik [10]. Data atau objek yang memiliki sifat atau atribut yang mirip akan dimasukkan ke dalam klaster yang sama, sedangkan yang memiliki perbedaan karakteristik akan ditempatkan pada klaster yang berbeda [11].

Tahapan algoritma *K-Means* antara lain:

- 1) Menentukan nilai k yang akan dibentuk (jumlah *cluster*)
- 2) Menentukan pusat klaster (*centroid*) dengan random, selanjutnya menghitung dengan rumus:

$$D(ij) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

Dengan:

$D(ij)$ = Jarak data ke i ke pusat cluster j

x_{ki} = Nilai atribut data ke i ke j

x_{kj} = Nilai atribut data ke j ke k

- 3) Menghitung jarak dari tiap data ke tiap centroid

$$d = \sum_{i=1}^n (x_i - y_i)^2$$

Dengan:

x_i = Nilai dari objek ke- i

y_i = Nilai dari *centroid* ke- i

n = Jumlah data dalam suatu klaster

- 4) Setiap data akan memilih centroid dengan jarak terdekat.

Kemudian, tentukan kembali posisi *centroid* dengan menghitung rata-rata dari seluruh data dalam satu klaster.

- 5) Ulangi proses ini (kembali ke langkah ketiga) apabila posisi *centroid* yang baru berbeda dengan yang sebelumnya.

Dengan menggunakan teknik ini, data akan dibagi yang akan dijadikan beberapa kelompok sehingga dapat ditemukan data serupa dengan kelompok lain. Konsep dasar algoritma *k-means* sangat mudah: minimalkan *Sum of Squared Error* (SSE) antara item data yang memiliki *k centroid*. Menemukan anggota setiap cluster dan menentukan posisi pusat *cluster* adalah dua tugas yang sebenarnya diselesaikan oleh algoritma *k-means*. Mencari data yang diklasterkan, X_{ij} ($i=1, \dots, n; j=1, \dots, m$), dimana n adalah jumlah dari banyak data yang akan diklaster dan m merupakan banyaknya variabel, adalah langkah pertama dalam proses *clustering*. Setiap pusat *cluster* dipilih secara acak pada awal iterasi [12].

3.2 Data dan Variabel

Penelitian ini menggunakan data kuantitatif berupa data akademik dan keaktifan yang diperoleh dari Pesantren Darul Arafah Raya. Data diperoleh dari Santri dan santriwati berjumlah 1010 data, yang terdiri dari 5 kriteria penilaian, yakni; data nilai akademik, data absensi, data sikap, riwayat prestasi, dan keaktifan ekstrakurikuler. Data berbentuk non numerik atau huruf perlu dikonversi lagi agar lebih mudah di proses dan berbentuk Numpy saat diinput ke sistem. Variable yang dijadikan acuan dalam penilain santri berprestasi dan tidak berprestasi antara lain ialah:

Tabel 1. Variabel Penelitian

No	Variabel	Keterangan Penilaian
1.	Nilai Akademik (p1)	Nilai rata-rata dengan range nilai 1-100
2.	Absensi	Konversi nilai non numerik A-D menjadi 1-100
3.	Riwayat Prestasi	Nilai rata-rata prestasi 1-100
4.	Organisasi/Ekstrakurikuler	Nilai 1-100
5.	Sikap	Konversi nilai non numerik A-D menjadi 1-100

Konversi Nilai non numerik untuk nilai sikap:

1. A = 85
2. B = 80
3. C = 75
4. D = 70

Konversi Nilai Non numerik untuk nilai absensi:

1. Tidak hadir 1–2 kali = A = 85
2. Tidak hadir 3 kali = B = 80
3. Tidak hadir 4 kali = C = 75
4. Tidak hadir 5 kali = D = 70

3.3 Metode Validasi Performa Algoritma

Pada penelitian ini digunakan metode *Silhouette Coefficient* sebagai metode validasi pengujian algoritma. *Silhouette Coefficient* adalah salah satu metode evaluasi internal yang digunakan untuk menilai seberapa baik objek ditempatkan dalam suatu klaster [13]. Metode ini memanfaatkan perbandingan antara rata-rata jarak objek terhadap anggota dalam klaster sendiri (*intra-cluster distance*) dan rata-rata jarak objek terhadap klaster terdekat lainnya (*inter-cluster distance*) [14].

Nilai *Silhouette Coefficient* berada dalam rentang -1 hingga 1:

1. Nilai yang mendekati 1 mengindikasikan bahwa objek berada dalam klaster yang tepat (sangat sesuai dengan klasternya).
2. Jika nilai mendekati 0, maka posisi objek cenderung berada di peralihan dua klaster.
3. Nilai negatif (< 0) menunjukkan objek lebih dekat dengan klaster lain daripada klasternya saat ini (penempatan kurang tepat).

Rumus yang digunakan Adalah:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

Keterangan:

1. $a(i)$ dihitung sebagai rata-rata jarak data ke- i terhadap data lain pada kluster yang sama (*intra-cluster distance*).
2. $b(i)$ = rata-rata jarak data i terhadap seluruh data pada kluster terdekat (*inter-cluster distance*).
3. $s(i)$ = nilai *Silhouette* untuk data i .

Nilai akhir *Silhouette* untuk seluruh data diperoleh dengan menghitung rata-rata semua nilai $s(i)$. dengan n adalah jumlah data keseluruhan.

$$S = \frac{1}{n} \sum_{i=1}^n s(i) \quad (3)$$

Prosedur dalam menghitung *Silhouette Coefficient* disajikan sebagai berikut:

1. Tentukan hasil klusterisasi menggunakan algoritma *K-Means*, di mana setiap data telah masuk ke dalam kluster tertentu.
2. Hitung $a(i)$ untuk setiap data i , yaitu rata-rata jarak antara data i dengan seluruh anggota dalam kluster sendiri.
3. Hitung $b(i)$ untuk setiap data i , yaitu rata-rata jarak antara data i dengan seluruh anggota dari kluster terdekat.
4. Hitung $s(i)$ untuk setiap data i menggunakan rumus *Silhouette Coefficient*.
5. Rata-ratakan nilai *Silhouette* seluruh data untuk mendapatkan nilai kualitas kluster secara keseluruhan.
6. Interpretasi hasil nilai mendekati 1 menunjukkan kualitas kluster baik, nilai mendekati 0 menunjukkan pemisahan kurang jelas, sedangkan nilai negatif menunjukkan adanya salah klasifikasi.

Metode *Silhouette Coefficient* digunakan sebagai pendekatan validasi untuk mengukur tingkat optimalitas hasil pengelompokan santri dan santriwati berprestasi yang diperoleh melalui algoritma *K-Means*, sehingga dapat dipastikan kesesuaian kluster dengan karakteristik data.

4. Hasil dan Pembahasan

4.1 Data Penelitian

Pengumpulan data yang menggunakan data akademik yang menjadi dasar analisis. Pendekatan ini memungkinkan peneliti memperoleh informasi yang akurat dan terpercaya langsung dari sumber utama, sehingga data yang dikumpulkan relevan dan mendukung pelaksanaan penelitian secara efektif. Data yang dikumpulkan diambil dari santri/santriwati kelas X-XII. Data penelitian terdapat pada tabel 2 berikut:

Tabel 2. Data Akademik Santri/Wati

No	NISN	Nama	Kelas	Jenis Kelamin	Nilai Akademik	Adab/Sikap	Nilai Ekskul	Prestasi	Absensi
1	81918269	A Rasyd Suryanta Singarimbun	X A	L	88	B	81	89	B
2	73899550	Alhafid Ibnu	X B	L	81	B	80	78	D
3	83968470	Danish Raziq Mulia	X C	L	83	C	84	85	B
4	21474836	Gilang Akbar Wijaya Sirait (Lq)	X D	L	78	B	83	73	B
5	84212218	Habib Ziqri Ananda Suherman	X E	L	80	C	83	88	C
6	98875815	Hazrullah Azwar	X F	L	84	D	85	70	A

No	NISN	Nama	Kelas	Jenis Kelamin	Nilai Akademik	Adab/Sikap	Nilai Ekskul	Prestasi	Absensi
7	95380166	M. Gathfan Liyandi	X G	L	83	C	70	84	B
8	89893360	Maulidin Siddik	X H	L	90	B	97	89	A
9	95133187	Teuku Raja Alkisah	X I	L	76	C	82	10 0	A
10	83268518	Arjuna	X J	L	80	B	98	80	C
...
1010	88981756	Zaskia Nur Fadilla Lumban Batu	7 A	P	88	B (8 0)	89	77	C (7 5)

4.2. Implementasi Algoritma K-Means

Implementasi *K-Means* dalam penelitian ini dimulai setelah proses pengumpulan data selesai dan menentukan nilai angka dari nilai yang bersifat non-numerik. Adapun tahapannya sebagai berikut:

1) Menentukan Jumlah *Cluster*

Tahap ini merupakan tahap untuk menentukan jumlah *cluster* yang akan digunakan.

Pada sistem *clustering* pengelompokan Santri/santriwati berprestasi akan menggunakan 2 *cluster* yaitu:

- a. *Cluster* Pertama (C1) = Berprestasi
- b. *Cluster* Kedua (C2) = Tidak Berprestasi

2) Menentukan Pusat *Cluster*

Pada tahap ini akan menentukan nilai pusat *Cluster (Centroid)* dari data yang digunakan. Berikut adalah *centroid* awal dari data penyakit yang akan digunakan.

Tabel 3. Centroid Awal

Centroid	F1	F2	F3	F4	F5	Cluster
1	92	85	96	69	75	Berprestasi
2	70	75	75	77	85	Tidak Berprestasi

3) Perhitungan Jarak Data dengan *Centroid*

Proses berikutnya melibatkan perhitungan jarak antara masing-masing data dengan masing-masing pusat *cluster* yang telah ditentukan sebelumnya menggunakan rumus *Euclidean Distance*. Dengan demikian, dapat diketahui data mana yang memiliki jarak terdekat ke *centroid*.

Kemudian untuk menghitung dari data ke *centroid* yang dimiliki:

$$D(i,j) = \sqrt{\sum_{k=1}^m (x_{ki} - x_{kj})^2}$$

Jarak data A Rasyd Suryanta Singarimbun (1) dengan centroid 1 dan 2:

$$D(i,j) = \sqrt{(88 - 92)^2 + (80 - 85)^2 + (81 - 96)^2 + (89 - 69)^2 + (80 - 75)^2} = 26.2869$$

$$D(i,j) = \sqrt{(88 - 70)^2 + (80 - 75)^2 + (81 - 75)^2 + (89 - 77)^2 + (80 - 85)^2} = 23.5372$$

Jarak data Alhafid Ibnu dengan (2) centroid 1 dan 2:

$$D(i, j) = \sqrt{(81 - 92)^2 + (80 - 85)^2 + (80 - 96)^2 + (78 - 69)^2 + (70 - 75)^2} = 22.5389$$

$$D(i, j) = \sqrt{(81 - 70)^2 + (80 - 75)^2 + (80 - 75)^2 + (78 - 77)^2 + (70 - 85)^2} = 19.9249$$

.....

Jarak data (data ke 1010) dengan centroid 1 dan 2:

$$D(i, j) = \sqrt{(88 - 92)^2 + (80 - 85)^2 + (89 - 96)^2 + (77 - 69)^2 + (75 - 75)^2} = 12.4097$$

$$D(i, j) = \sqrt{(88 - 70)^2 + (80 - 75)^2 + (89 - 75)^2 + (88 - 77)^2 + (75 - 85)^2} = 25.3968$$

Hasil perhitungan jarak data santri pada perhitungan iterasi ke -1 dapat dilihat tabel berikut:

Tabel 4. Hasil Perhitungan Jarak Data Dengan Centroid Awal Iterasi 1

No	Nama Santri	C1	C2	Cluster
1.	A Rasyd Suryanta Singarimbun	26.2869	23.5372	C2
2.	Alhafid Ibnu	21.9773	16.4924	C2
3.	Danish Raziq Mulia	24.6171	18.4120	C2
4.	Gilang Akbar Wijaya Sirait	20.7605	13.9284	C2
5.	Habib Ziqri Ananda Suherman	27.8209	19.6214	C2
6.	Hazrullah Azwar	22.6053	19.2354	C2
7.	M. Gathfan Liyandi	33.2716	16.3707	C2
8.	Maulidin Siddik	23.0217	32.4500	C1
9.	Teuku Raja Alkisah	40.1622	24.7790	C2
10.	Arjuna	17.1464	27.6225	C1
...
1010	Zaskia Nur Fadilla Lumban Batu	30.9354	18.3303	C2

4) Hasil *Clustering*

Tabel 5. Hasil *Clustering*

Cluster	Nama Santri	Kategori
Cluster 2	A Rasyd Suryanta Singarimbun	Tidak Berprestasi
Cluster 1	Alhafid Ibnu	Berprestasi
Cluster 2	Danish Raziq Mulia	Tidak Berprestasi
Cluster 2	Gilang Akbar Wijaya Sirait	Berprestasi
Cluster 2	Habib Ziqri Ananda Suherman	Tidak Berprestasi
Cluster 2	Hazrullah Azwar	Berprestasi
Cluster 2	M. Gathfan Liyandi	Tidak Berprestasi
Cluster 1	Maulidin Siddik	Tidak Berprestasi
Cluster 2	Teuku Raja Alkisah	Tidak Berprestasi
Cluster 1	Arjuna	Berprestasi
...
Cluster 2	Zaskia Nur Fadilla Lumban Batu	Berprestasi

Tabel 5 mengelompokkan beberapa nama santri/wati ke dalam dua cluster berdasarkan penilaian lima variabelnya, antara lain nilai akademik, nilai absensi, nilai Riwayat prestasi, ekstrakurikuler, dan sikap. *Cluster* 1 menjelaskan berprestasi, sedangkan *Cluster* 2 tidak berprestasi. Dari 1010 data santri, 490 diantaranya dikategorikan ke dalam C1 (Berprestasi) dan 520 santri masuk ke cluster C2 (Tidak Berprestasi).

4.3 Validasi Pengujian Performa Algoritma

Setelah proses klusterisasi dengan algoritma *K-Means* menghasilkan 2 klaster, yaitu *Cluster* 1 (Santri dan santriwati Berprestasi) dengan 490 anggota dan *Cluster* 2 (Santri dan santriwati Tidak Berprestasi) dengan 520 anggota, Evaluasi kualitas klaster dilakukan melalui metode *Silhouette Coefficient*.

4.3.1 Hasil Perhitungan *Silhouette Coefficient*

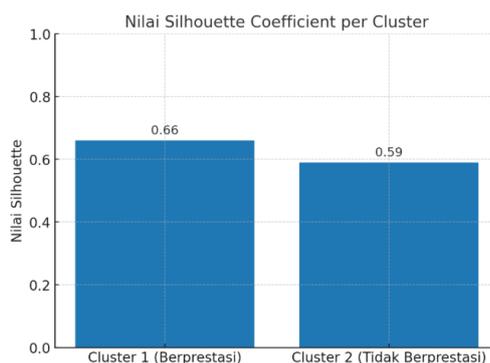
Perhitungan dilakukan dengan menghitung nilai rata-rata kedekatan antar data dalam kluster (*intra-cluster distance, a(i)*) dan jarak data ke kluster terdekat (*inter-cluster distance, b(i)*). Nilai *Silhouette* setiap data kemudian dirata-ratakan untuk memperoleh kualitas keseluruhan klusterisasi. Hasil pengujian didapatkan yakni:

Tabel 6. Hasil pengujian Algoritma

Klaster	Jumlah Anggota	Nilai Rata-rata <i>Silhouette</i>
Cluster 1 (Berprestasi)	490	0.66
Cluster 2 (Tidak Berprestasi)	520	0.59
Keseluruhan	1.010	0.62

Interpretasi Hasil:

1. Nilai rata-rata *Silhouette Coefficient* keseluruhan adalah 0.62, yang berada pada kategori cukup baik.
2. *Cluster 1* (Berprestasi) memiliki nilai *Silhouette* lebih tinggi (0.66), menunjukkan bahwa pengelompokan data pada kluster ini lebih konsisten dan homogen.
3. *Cluster 2* (Tidak Berprestasi) memiliki nilai 0.59, yang berarti masih terdapat sebagian data yang posisinya relatif dekat dengan kluster lain, namun secara umum masih valid.
4. Dengan demikian, algoritma *K-Means* mampu secara efektif mengelompokkan 1.010 santri dan santriwati menjadi dua kategori berdasarkan variabel nilai akademik, absensi, sikap, keterlibatan organisasi/ekstrakurikuler, dan riwayat prestasi, serta menghasilkan kualitas kluster yang cukup baik menurut evaluasi *Silhouette*.



Gambar 2. Grafik Batang nilai *Silhouette Coefficient*

Berikut adalah grafik batang yang menggambarkan nilai *Silhouette Coefficient* pada masing-masing *cluster*. *Cluster* Berprestasi (0.66) dan *Cluster* Tidak Berprestasi (0.59).

4.4. Antarmuka Sistem

1) Halaman Data *Centroid*

Halaman ini merupakan tampilan halaman dari data kluster, pengelompokan santri didasarkan pada nilai atribut yang ada pada santri. *Centroid* diambil secara acak, dan hasil pengelompokan diambil dari beberapa iterasi. Berikut tampilan dari penilaian iterasi pertama.

#	F1	F2	F3	F4	F5
C1	92.00	85.00	96.00	69.00	75.00
C2	70.00	75.00	75.00	77.00	85.00

Gambar 2. Centroid Awal untuk iterasi ke-1

Cluster Setiap Data:

Show 10 entries Search:

DATA	CLUSTER
DA RASYD SURYANTA SINGARIMBUN	C2
DA'inul Fitriya R. Al-Mujahidah	C2
DABDI GHOFFARI	C2
DABDILLAH HASIBUAN	C1
DABDUL FATHAN HALIMSYAH RAMBE	C2
DABDUL GHANI	C2
DABDUL RAHMAN	C2
DABDURRAHIM WINATA	C2
DABDURRAHMAN AHMAD AL FAQIH	C1
DABID AL-MALIK SEMBIRING	C2

Showing 1 to 10 of 1,010 entries Previous 1 2 3 4 5 ... 101 Next

Gambar 3. Hasil *Cluster* data Awal untuk iterasi ke-1

2) Halaman Data Hasil Pengelompokan

Ini merupakan hasil pengelompokan data santri, setelah beberapa iterasi hingga terhenti di iterasi 35 dan pusat *centroid* telah dinyatakan konvergen, hasil akan terlihat santri mana yang dikelompokkan berprestasi dan tidak. Data menunjukkan santri berprestasi sebanyak 490, dan tidak berprestasi sebanyak 520 santri.

#	NAMA SISWA	CLUSTER	KETERANGAN
1	A RASYD SURYANTA SINGARIMBUN	C1	Tidak Berprestasi
2	ALHAFID IBNU	C2	Berprestasi
3	DANISH RAZIQ MULIA	C1	Tidak Berprestasi
4	GILANG AKBAR WIJAYA SIRAIT (LQ)	C2	Berprestasi
5	HABIB ZIQRI ANANDA SUHERMAN	C1	Tidak Berprestasi
6	HAZRULLAH AZWAR	C2	Berprestasi
7	M. GATHFAN LIYANDI	C1	Tidak Berprestasi
8	MAULIDIN SIDDIK	C1	Tidak Berprestasi
9	TEUKU RAJA ALKISAH	C1	Tidak Berprestasi
10	ARJUNA	C2	Berprestasi
Total Santri Berprestasi:			490
Total Santri Tidak Berprestasi:			520

Gambar 4. Halaman Data Hasil Pengelompokan

Centroid Baru:

Show 10 entries Search:

#	F1	F2	F3	F4	F5
C1	83.77	77.57	80.59	91.43	77.01
C2	80.97	77.58	83.61	73.28	77.72

Showing 1 to 2 of 2 entries Previous 1 Next

Gambar 5. *Centroid* baru untuk iterasi terakhir ke-35

Pada Gambar 5, dapat dilihat bahwa proses iterasi sudah berhenti karena posisi *centroid* tidak lagi berubah secara signifikan, atau pembagian kluster terhadap data sudah stabil.

4.4 Pembahasan

Hasil penerapan *K-Means Clustering* terhadap 1.010 data santri dan santriwati di Pesantren Darul Arafah Raya menghasilkan dua kelompok utama, yaitu santri berprestasi (490 orang) dan tidak berprestasi (520 orang), dengan proses iterasi berhenti pada langkah ke-35 saat *centroid* mencapai konvergensi. Analisis *centroid* akhir memperlihatkan bahwa santri dalam *Cluster* 1 memiliki capaian lebih tinggi pada variabel akademik (83,77), riwayat prestasi (80,59), dan sikap (91,43), sehingga dikategorikan sebagai kelompok berprestasi, sementara *Cluster* 2

menunjukkan kelemahan terutama pada variabel sikap (73,28) dan ekstrakurikuler (77,72). Evaluasi kualitas kluster menggunakan *Silhouette Coefficient* memperkuat hasil tersebut, dengan nilai rata-rata keseluruhan 0,62 yang berada pada kategori cukup baik. *Cluster 1* memiliki nilai lebih tinggi (0,66), menandakan pengelompokan yang konsisten dan homogen, sedangkan *Cluster 2* memperoleh nilai 0,59, yang meskipun menunjukkan sebagian data relatif dekat dengan kluster lain, tetap valid sebagai hasil dari kelompok santri tidak berprestasi. Dengan demikian, penelitian ini mengindikasikan bahwa metode *K-Means* tidak hanya mampu membentuk dua kategori yang sesuai dengan kondisi riil, tetapi juga menghasilkan kualitas kluster yang memadai untuk mendukung pengambilan keputusan dalam pembinaan santri/wati.

Temuan tersebut mendukung hasil penelitian sebelumnya yang dilakukan oleh [15], yang juga menggunakan algoritma *K-Means* dalam pengelompokan mahasiswa berdasarkan nilai akademik dan absensi, di mana diperoleh nilai *Silhouette Score* yakni 0.61. Hasil tersebut menunjukkan bahwa *K-Means* mampu menghasilkan pemisahan kluster yang cukup baik pada data pendidikan. Penelitian lain oleh [16] menunjukkan bahwa algoritma *K-Means* lebih efektif dalam membentuk kluster yang rapat dibandingkan dengan *Hierarchical Clustering*, terutama ketika digunakan untuk mengelompokkan kinerja siswa di sekolah menengah. Penelitian ini memperkuat hasil tersebut dengan menunjukkan nilai *Silhouette* yang lebih tinggi pada kelompok berprestasi, menandakan bahwa santri dengan capaian akademik dan non-akademik baik memiliki pola yang lebih konsisten dalam pengelompokannya.

Dalam hal kontribusi, penelitian ini memperkuat penelitian sebelumnya dengan menawarkan konteks baru, yaitu pengelompokan santri dan santriwati di lingkungan pesantren, yang hingga saat ini masih jarang diteliti dalam literatur data mining. Penemuan ini memperluas penggunaan *K-Means* dari sektor pendidikan formal ke pendidikan berbasis pesantren, sekaligus menonjolkan pentingnya variabel multidimensional (akademik, absensi, sikap, ekstrakurikuler, dan prestasi) untuk menyajikan gambaran menyeluruh tentang prestasi siswa. Oleh karena itu, studi ini tidak hanya menguatkan temuan-temuan sebelumnya, tetapi juga memperluas aplikasi algoritma *K-Means* dalam analisis data pendidikan di berbagai konteks institusional

5. Simpulan

Penelitian ini berhasil merancang dan mengimplementasikan sistem berbasis web untuk pemilihan santri dan santriwati berprestasi menggunakan algoritma *K-Means Clustering*. Hasil pengelompokan menunjukkan bahwa dari total 1.010 data santri dan santriwati, terbentuk dua kluster utama, yaitu kluster Berprestasi sebanyak 490 santri dan santriwati dan kluster Tidak Berprestasi sebanyak 520 santri dan santriwati. Evaluasi kualitas kluster menggunakan *Silhouette Coefficient* menghasilkan nilai rata-rata keseluruhan sebesar 0,62, yang termasuk kategori cukup baik. Secara rinci, kluster Berprestasi memperoleh nilai 0,66, menandakan bahwa data dalam kluster ini relatif konsisten dan homogen, sedangkan kluster Tidak Berprestasi memperoleh nilai 0,59, yang menunjukkan masih terdapat sebagian data yang posisinya relatif dekat dengan kluster lain, namun secara umum masih valid. Dengan demikian, algoritma *K-Means* terbukti mampu mengelompokkan santri dan santriwati berdasarkan variabel nilai akademik, absensi, sikap, keterlibatan berorganisasi/ekstrakurikuler, dan riwayat prestasi, serta memberikan hasil pengelompokan dengan kualitas yang cukup baik untuk mendukung pengambilan keputusan dalam konteks pendidikan pesantren.

Daftar Referensi

- [1] G. Ayu, A. Ikhwan, and R. A. Putri, "Sistem Informasi Manajemen Aset Tetap Menggunakan Metode Waterfall," *RESOLUSI : Rekayasa Teknik Informatika dan Informatika*, vol. 3, no. 6, pp. 187–296, 2023, doi: <https://doi.org/10.30865/resolusi.v3i6.992>.
- [2] R. Riadi and Mesran, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Analisa Penjualan Parfume," *Journal of Informatics, Electrical and Electronics Engineering*, vol. 2, no. 4, pp. 138–145, Jun. 2023, doi: [10.47065/jieeee.v2i4.1181](https://doi.org/10.47065/jieeee.v2i4.1181).
- [3] G. Sonia and R. A. Putri, "Penerapan Metode K-Means Clustering Untuk Mengelompokkan Data Kelayakan Penerima Bantuan Renovasi Rumah," *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 2, pp. 442–455, Sep. 2023, doi: [10.47065/bits.v5i2.4298](https://doi.org/10.47065/bits.v5i2.4298).
- [4] J. Hutagalung, Y. H. Syahputra, and Z. P. Tanjung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 1, pp. 606–620, Mar. 2022, doi: <https://dx.doi.org/10.35957/jatisi.v9i1.1516>.

-
- [5] S. Kurniawan, A. M. Siregar, and H. Y. Novita, "Penerapan Algoritma K-Means dan Fuzzy C-Means Dalam Mengelompokan Prestasi Siswa Berdasarkan Nilai Akademik," *Scientific Student Journal for Information, Technology and Science*, vol. IV, no. 1, pp. 73–81, Jan. 2023.
- [6] Y. Amri, "Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik," *MENARA Ilmu*, vol. 15, pp. 143–152, Oct. 2021, doi: <http://dx.doi.org/10.31869/mi.v15i2.3180>.
- [7] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *Journal of Information Systems and Informatics*, vol. 3, no. 3, pp. 424–439, Sep. 2021, doi: <https://doi.org/10.51519/journalisi.v3i3.164>.
- [8] R. Putra Primanda, A. Alwi, and D. Mustikasari, "Data Mining Seleksi Siswa Berprestasi Untuk Menentukan Kelas Unggulan Menggunakan Metode K-Means Clustering (Studi Kasus di MTS Darul Fikri)," Ponorogo, Apr. 2021.
- [9] D. Apriandi, R. M. Sari, and M. I. Sarif, "Analisis Clustering Untuk Menentukan Siswa Berprestasi di SMK Swasta TI Panca Dharma Stungkit Menggunakan Metode K-Means," *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 1117–1129, Aug. 2024, doi: [10.33395/jmp.v13i1.13959](https://doi.org/10.33395/jmp.v13i1.13959).
- [10] F. Pramataning Dewi, P. Siwi Aryni, and Y. Umidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," MEI, 2011.
- [11] R. Alhapizi, M. Nasir, and I. Effendy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," *Journal of Software Engineering Ampera*, vol. 1, no. 1, pp. 1-14. 2020. [Online]. Available: <https://journal-computing.org/index.php/journal-sea/index>
- [12] U. S. Hanifatus, E. H. Wahyu Dyah, N. Fitriyah, and C. Saleh, "Implementasi Pengelompokan Kelas Berdasarkan Kemampuan Akademik di MI Mambaul Ma'arif Denanyar Jombang," *PENDAGOGIA: Jurnal Pendidikan Dasar*, vol. 2, no. 2, pp. 137-146, 2022.
- [13] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Ilmiah Informatika Dan Ilmu Komputer*, vol. 3, no. 1, pp. 46–56, Mar. 2024, doi: [10.58602/jima-ilkom.v3i1.26](https://doi.org/10.58602/jima-ilkom.v3i1.26).
- [14] K. S. Nadhiva, A. Triayudi, and E. T. E. Handayani, "Implementasi Sistem Informasi Rekam Medis Berbasis Web Klinik Gigi menggunakan Metode Waterfall dan PIECES Framework," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 10, no. 1, pp. 168-177, 2022, doi: [10.26418/justin.v10i1.50997](https://doi.org/10.26418/justin.v10i1.50997).
- [15] S. Haviyola, Susilawati, and M. Jajuli, "Pengelompokan Prestasi Siswa Guna Kualifikasi Beasiswa Berdasarkan Data Nilai Menggunakan Algoritma K-Means," *JATI*, vol. 7, no. 4, pp. 2786-2791, 2023. doi: <http://dx.doi.org/10.36040/jati.v7i4.7200>.
- [16] A. Yudistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 1, no. 1, pp. 20–28, Mar. 2023, doi: [10.58602/jaiti.v1i1.22](https://doi.org/10.58602/jaiti.v1i1.22).